

Asymptotics of Fixed Point Distributions for Inexact Monte Carlo Algorithms

M. A. Clark*

*Center for Computational Sciences, Boston University,
3 Cummington Street, Boston, MA 02215, United States of America*

A. D. Kennedy†

*School of Physics, The University of Edinburgh,
Mayfield Road, Edinburgh, EH9 3JZ, United Kingdom*

(Dated: February 1, 2008)

We introduce a simple general method for finding the equilibrium distribution for a class of widely used inexact Markov Chain Monte Carlo algorithms. The explicit error due to the non-commutivity of the updating operators when numerically integrating Hamilton's equations can be derived using the Baker–Campbell–Hausdorff formula. This error is manifest in the conservation of a “shadow” Hamiltonian that lies close to the desired Hamiltonian. The fixed point distribution of inexact Hybrid algorithms may then be derived taking into account that the fixed point of the momentum heatbath and that of the molecular dynamics do not coincide exactly. We perform this derivation for various inexact algorithms used for lattice QCD calculations.

PACS numbers: 02.50.Tt, 02.70.Ns, 11.15.Ha, 45.20.Jj

I. INTRODUCTION

The first algorithms used to simulate lattice QCD with dynamical quarks fell into three classes: Monte Carlo [1, 2], Langevin [3], and Molecular Dynamics (microcanonical) [4, 5] methods. The first two involve taking many essentially independent small random steps, leading to a dynamical critical exponent $z = 2$ typical of Brownian motion, where this exponent relates the autocorrelation time of the Markov process to the correlation length of the physical system. The steps are small either in the sense of making only a small global change in the gauge field or by making a larger change locally, such as updating a single link variable. The third suffers from the problem of not being ergodic in general, and therefore not necessarily generating the correct distribution of configurations; nevertheless, although the updates are also built out of many small steps these steps are correlated so as to lead to a dynamical critical exponent $z = 1$, and it can thus be considered as a “large step” algorithm. The last two algorithms also depend upon a step size parameter $\delta\tau$, and are only “exact” in the limit $\delta\tau \rightarrow 0$.

It was then realized [6] that by combining Molecular Dynamics with a momentum refreshment heatbath an ergodic large step “Hybrid” algorithm resulted, although it still suffered from step size errors in the equilibrium distribution. Shortly afterwards it was found that by combining such Hybrid updates with a Metropolis Monte Carlo acceptance test these step size errors could be completely eliminated [7]: this is called the Hybrid Monte Carlo algorithm. Both the idea of combining Molecular Dynamics with momentum refreshment [8] and that of correcting the Langevin algorithm by a Metropolis test [9] had been introduced previously in other fields.

The introduction of the Hybrid Monte Carlo algorithm did not end the use of inexact algorithms for two reasons: firstly some people believe that the volume dependence of the cost of the Hybrid Monte Carlo algorithm outweighs the advantage of its exactness, and secondly they wanted to simulate with what are now called “rooted staggered quarks” for which there is no explicit local form for the action, but instead the square- or fourth-root of the corresponding determinant is required. A way of doing this with step size errors of $O(\delta\tau^2)$ was introduced with the R algorithm [10]. Only recently has the RHMC algorithm [11, 12] provided an efficient exact alternative.

The purpose of the present paper is to analyze the inexact algorithms mentioned above, as well as several other interesting variants. The approach we take is new, namely we view all the algorithms as a combination of a Molecular Dynamics trajectory with a momentum refreshment; from this point of view the Langevin algorithm is just a single step Molecular Dynamics trajectory. Although in principle, any numerical integration scheme could be used for the Molecular Dynamics integration in practice all the algorithms use symmetric symplectic integrators (or closely related integrators). Such integrators have several remarkable properties, such as being reversible and area preserving, which

*Electronic address: mikec@bu.edu

†Electronic address: adk@ph.ed.ac.uk

are required for the exact Hybrid Monte Carlo algorithm. We make use of the remarkable property that all symplectic integrators exactly conserve a “shadow Hamiltonian” close to the desired one for small enough step sizes to produce a uniform and simple asymptotic expansion for the equilibrium (fixed point) distribution of the corresponding inexact Markov chains.

We then turn to the class of inexact algorithms that use a noisy estimate of the force due to the fermionic determinant. We introduce the pseudofermionic χ algorithm as an aid to establishing that the R algorithm, which may be considered as an interpolation between the χ and R_0 algorithms, has only $O(\delta\tau^2)$ errors in its equilibrium distribution. We also establish that one of our original motivations — namely trying to find an variant of the R algorithm with errors falling as a higher power of the step size — is almost certainly doomed to failure as the unwanted noise contributions to the error are necessarily positive.

This paper is organised as follows: in §II we use the Baker–Campbell–Hausdorff (BCH) formula to derive the form of the “shadow” Hamiltonian that is conserved by various symplectic integrators. A brief description of Hybrid stochastic algorithms is given in §III, and the resulting fixed points of these algorithms are derived in §IV. We then turn our attention to noisy algorithms, specifically as applied to fermion theories, in §V. For the reader’s convenience we collect as appendices derivations of some important results that are rather difficult to find in an accessible form in the literature.

II. SYMPLECTIC INTEGRATORS

We are interested in finding the classical trajectory in phase space of a system described by a Hamiltonian $H(q, p) = T(p) + S(q) = \frac{1}{2}p^2 + S(q)$. It has been known for a long time that the leapfrog integration scheme has many desirable properties, and higher-order generalizations have been discovered several times in different fields [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. The basic idea of such a *symplectic integrator* is to write the time evolution operator as

$$\begin{aligned} \exp\left(\tau \frac{d}{dt}\right) &\equiv \exp\left(\tau \left\{ \frac{\partial p}{\partial t} \frac{\partial}{\partial p} + \frac{\partial q}{\partial t} \frac{\partial}{\partial q} \right\}\right) \\ &= \exp\left(\tau \left\{ -\frac{\partial H}{\partial q} \frac{\partial}{\partial p} + \frac{\partial H}{\partial p} \frac{\partial}{\partial q} \right\}\right) \equiv e^{\tau \hat{H}} \\ &= \exp\left(\tau \left\{ -S'(q) \frac{\partial}{\partial p} + T'(p) \frac{\partial}{\partial q} \right\}\right), \end{aligned}$$

where \hat{H} is the *Hamiltonian vector field*. In the second step we have made use of Hamilton’s equations, and thus implicitly the fundamental 2-form¹ $\omega = dq \wedge dp$.

From the structure of our Hamiltonian, namely the fact that the kinetic energy T is a function only of p and the potential energy S is a function only of q , it follows that $\exp\left[-\tau S'(q) \frac{\partial}{\partial p}\right]$ and $\exp\left[\tau T'(p) \frac{\partial}{\partial q}\right]$ are trivial² to evaluate. Let us write

$$Q \equiv T'(p) \frac{\partial}{\partial q} \quad \text{and} \quad P \equiv -S'(q) \frac{\partial}{\partial p}, \tag{1}$$

so that by Taylor’s theorem

$$\begin{aligned} e^{tQ} : f(q, p) &\mapsto f(q + tT'(p), p), \\ e^{tP} : f(q, p) &\mapsto f(q, p - tS'(q)); \end{aligned} \tag{2}$$

¹ For simplicity of presentation we shall consider here only the case where the fundamental symplectic 2-form is $\omega = dq \wedge dp$. For gauge theories we need to use $\omega = -\sum_i d(p_i \theta_i)$ where the θ_i are the left-invariant Maurer–Cartan forms on a Lie group manifold [24], but this generalization is straightforward.

² For gauge fields the p lie in a Lie algebra and the q in the corresponding Lie group, so the evaluation is straightforward if not entirely trivial.

then from the BCH formula (B3) we find that the QP symplectic integrator leads to the time evolution given by

$$\begin{aligned}
& (e^{\delta\tau Q} e^{\delta\tau P})^{\tau/\delta\tau} \\
&= \left(\exp \left[(Q + P)\delta\tau + \frac{1}{2}[Q, P]\delta\tau^2 + O(\delta\tau^3) \right] \right)^{\tau/\delta\tau} \\
&= \exp \left[\tau \left((Q + P) + \frac{1}{2}[Q, P]\delta\tau + O(\delta\tau^2) \right) \right] \\
&\equiv e^{\tau \widehat{H}'} = e^{\tau(Q+P)} + O(\delta\tau).
\end{aligned}$$

In addition to conserving energy to $O(\delta\tau)$ such symplectic integrators are manifestly area preserving.

The BCH formula tells us more than that there is an area-preserving approximate integration scheme for Hamilton's equations: it tells us that for each symplectic integrator there exists a "shadow" Hamiltonian H' close to the original one, which is *exactly conserved*. For the QP integrator the actual trajectories through phase space are integral curves of the vector field³

$$\begin{aligned}
\widehat{H}' &= Q + P + \frac{1}{2}[Q, P]\delta\tau + \frac{1}{12} \left\{ [Q, [Q, P]] - [P, [Q, P]] \right\} \delta\tau^2 - \frac{1}{24} [P, [Q, [Q, P]]] \delta\tau^3 \\
&\quad + \frac{1}{720} \left\{ -4 [P, [Q, [Q, [Q, P]]]] - 6 [[Q, P], [Q, [Q, P]]] + 4 [P, [P, [Q, [Q, P]]]] \right. \\
&\quad \left. - 2 [[Q, P], [P, [Q, P]]] - [Q, [Q, [Q, [Q, P]]]] + [P, [P, [P, [Q, P]]]] \right\} \delta\tau^4 + O(\delta\tau^5) \\
&= \left\{ p \frac{\partial}{\partial q} - S' \frac{\partial}{\partial p} \right\} + \frac{1}{2} \left\{ S' \frac{\partial}{\partial q} - p S'' \frac{\partial}{\partial p} \right\} \delta\tau + \frac{1}{12} \left\{ 2p S'' \frac{\partial}{\partial q} - (p^2 S''' + 2S' S'') \frac{\partial}{\partial p} \right\} \delta\tau^2 \\
&\quad + \frac{1}{12} \left\{ S' S'' \frac{\partial}{\partial q} - p(S' S''' + S''^2) \frac{\partial}{\partial p} \right\} \delta\tau^3 + \frac{1}{720} \left\{ 4 \left(-p^3 S^{(4)} + p(6S''^2 + 3S' S''') \right) \frac{\partial}{\partial q} \right. \\
&\quad \left. + \left(p^4 S^{(5)} - p^2(30S''' S'' + 6S^{(4)} S') - 12S'(2S''^2 + S' S''') \right) \frac{\partial}{\partial p} \right\} \delta\tau^4 + O(\delta\tau^5)
\end{aligned}$$

from equation (B3). In Appendix A we show that the vector field \widehat{H}' is a Hamiltonian vector field, that is there exists a Hamiltonian H' such that

$$\widehat{H}' = \frac{\partial H'}{\partial p} \frac{\partial}{\partial q} - \frac{\partial H'}{\partial q} \frac{\partial}{\partial p},$$

and that H' may be obtained explicitly by replacing the commutators in the BCH expansion (B3) of $\ln(e^T e^S)$ by Poisson brackets⁴ of the form

$$\{A, B\} \equiv \frac{\partial A}{\partial p} \frac{\partial B}{\partial q} - \frac{\partial A}{\partial q} \frac{\partial B}{\partial p};$$

so

$$\begin{aligned}
H' &= H + \frac{1}{2}\{T, S\}\delta\tau + \frac{1}{12} \left(\{T, \{T, S\}\} - \{S, \{T, S\}\} \right) \delta\tau^2 - \frac{1}{24} \{S, \{T, \{T, S\}\}\} \delta\tau^3 \\
&\quad + \frac{1}{720} \left(-4 \{S, \{T, \{T, \{T, S\}\}\}\} - 6 \{\{T, S\}, \{T, \{T, S\}\}\} \right. \\
&\quad \left. + 4 \{S, \{S, \{T, \{T, S\}\}\}\} - 2 \{\{T, S\}, \{S, \{T, S\}\}\} \right. \\
&\quad \left. - \{T, \{T, \{T, \{T, S\}\}\}\} + \{S, \{S, \{S, \{T, S\}\}\}\} \right) \delta\tau^4 + O(\delta\tau^5) \\
&= H + \frac{1}{2} p S' \delta\tau + \frac{1}{12} (p^2 S'' + S'^2) \delta\tau^2 + \frac{1}{12} p S' S'' \delta\tau^3 \\
&\quad + \frac{1}{720} \left(-p^4 S^{(4)} + p^2(6S' S''' + 12S''^2) + 12S'^2 S'' \right) \delta\tau^4 + O(\delta\tau^5).
\end{aligned}$$

³ Observe that H' is linear in $\frac{\partial}{\partial q}$ and $\frac{\partial}{\partial p}$ because all but that first term in the BCH expansion are commutators.

⁴ The Poisson brackets define a Lie algebra since they are manifestly antisymmetric and satisfy the Jacobi relation $\{A, \{B, C\}\} + \{B, \{C, A\}\} + \{C, \{A, B\}\} = 0$, q.v., Appendix A.

Note that H' cannot be written as the sum of a p -dependent kinetic term and a q -dependent potential term. This means that any hope that one could exactly “invert” this relation to find a Hamiltonian whose approximate integral exactly conserves H is forlorn.

A. Symmetric Integrators

It is immediately apparent that we can do better than this by using a symmetric symplectic integrator (B4) which just gives the familiar PQP leapfrog scheme for which

$$\begin{aligned} U_0(\delta\tau)^{\tau/dt} &= (e^{\delta\tau P/2} e^{\delta\tau Q} e^{\delta\tau P/2})^{\tau/\delta\tau} \\ &= \left(\exp \left[(P+Q) \delta\tau - \frac{1}{24} \left([P, [P, Q]] + 2[Q, [P, Q]] \right) \delta\tau^3 + O(\delta\tau^5) \right] \right)^{\tau/\delta\tau} \\ &= \exp \left[\tau \left((P+Q) - \frac{1}{24} \left([P, [P, Q]] + 2[Q, [P, Q]] \right) \delta\tau^2 + O(\delta\tau^4) \right) \right] \\ &= e^{\tau(P+Q)} + O(\delta\tau^2). \end{aligned}$$

The PQP integrator exactly conserves the Hamiltonian

$$H' = H + \frac{1}{24} \left\{ 2p^2 S'' - S'^2 \right\} \delta\tau^2 + \frac{1}{720} \left\{ -p^4 S^{(4)} + 6p^2 (S' S''' + 2S''^2) - 3S'^2 S'' \right\} \delta\tau^4 + O(\delta\tau^6),$$

whereas the QPQ integrator conserves

$$H' = H + \frac{1}{24} \left\{ -p^2 S'' + 2S'^2 \right\} \delta\tau^2 + \frac{1}{5760} \left\{ 7p^4 S^{(4)} - 24p^2 (3S' S''' + S''^2) + 96S'^2 S'' \right\} \delta\tau^4 + O(\delta\tau^6).$$

It is possible to construct symplectic integrators of arbitrarily high order; some aspects of this theory are discussed in Appendix C.

III. HYBRID STOCHASTIC ALGORITHMS

A. The Hybrid algorithm

The Hybrid algorithm [6, 25, 26], so called because it is a hybrid of the Langevin [3] and Molecular Dynamics algorithms, constructs a Markov process on a “fictitious” phase space consisting of the field variables of interest (the coordinates) and a set of corresponding “fictitious” momenta. It should be emphasized that these momenta have nothing to do with the momenta which occur in the field equations of motion and canonical quantization relations; they are just introduced to define a classical dynamics of the fields in a new “fictitious” time dimension. If we are considering a four dimensional field theory, for example, then our “fictitious” dynamics takes place in a new fifth time dimension. To this end we introduce a Hamiltonian $H(q, p)$ and a fundamental symplectic 2-form as in section II. We then iterate three Markov steps, each of which has the distribution e^{-H} as an approximate fixed point and which, when taken together, are ergodic.⁵

- The first such step is *momentum refreshment* whereby the fictitious momenta are chosen from a Gaussian heatbath.⁶
- The second step is to integrate Hamilton’s equations using an approximate integrator for some length of fictitious time τ , and then to flip the momenta

$$\begin{pmatrix} q \\ p \end{pmatrix} \mapsto U(\tau) \begin{pmatrix} q \\ p \end{pmatrix} \equiv \begin{pmatrix} q'' \\ p'' \end{pmatrix} \mapsto \begin{pmatrix} q'' \\ -p'' \end{pmatrix}.$$

It has the distribution $e^{-H'}$ as an exact fixed point, and it thus approximately preserves the desired Hamiltonian H . If the integrator is symplectic then this step is area-preserving, and if it is also symmetric then the step is reversible (this is why we incorporate the final momentum flip).

⁵ In some cases we should choose trajectory lengths from some random distribution for this to be true.

⁶ A generalization of this is the “partial momentum refreshment” [27] step used in the second-order Langevin [28, 29] or Kramer’s algorithm [30, 31, 32].

- The third step is a *momentum flip*, $F : p \mapsto -p$, which obviously preserves the Gaussian distribution of momenta. Since this step is immediately followed by a momentum refreshment it can be ignored in all cases but Kramer's algorithm [28, 29, 30, 31, 32].

B. Langevin Algorithm

The Langevin equation [3, 33] is just a special case of the Hybrid algorithm for a single leapfrog integration step with PQP ordering: in other words we use the symplectic integrator $e^{\delta\tau P/2} e^{\delta\tau Q} e^{\delta\tau P/2}$ using the notation of equation (1). This operator maps $(q, p) \mapsto (q', p')$ where $q' = q + p\delta\tau - \frac{1}{2}S'(q)\delta\tau^2$, which is the familiar Langevin equation when the the Gaussian distributed momenta p are written as Gaussian noise η and the time step as $\epsilon \equiv \delta\tau^2$:

$$q' = q - \frac{1}{2}S'(q)\epsilon + \eta\sqrt{\epsilon}.$$

The PQP symplectic integrator has the amusing property of giving rise to a simple closed form Langevin equation, but there are many other variants too. For example, the QPQ symplectic integrator $e^{\delta\tau Q/2} e^{\delta\tau P} e^{\delta\tau Q/2}$ leads to the equation

$$q' = q + p\delta\tau - \frac{1}{2}S'\delta\tau^2 - \frac{1}{4}S''p\delta\tau^3 - \frac{1}{16}S'''p^2\delta\tau^4 - \frac{1}{96}S^{(4)}p^3\delta\tau^5 - \frac{1}{768}S^{(5)}p^4\delta\tau^6 + O(\delta\tau^7), \quad (3)$$

the lowest-order PQP Campostrini wiggle (q.v., Appendix C) gives

$$\begin{aligned} q' = & q + p\delta\tau - \frac{1}{2}S'\delta\tau^2 - \frac{1}{6}S''p\delta\tau^3 + \frac{1}{24}[S''S' - S'''p^2]\delta\tau^4 \\ & + \frac{1}{432}\left[\left(6(3 + \sqrt[3]{2} + \sqrt[3]{4})S'''S' - 3(4 + 4\sqrt[3]{2} + 3\sqrt[3]{4})S''^2\right)p \right. \\ & \left. - (3 + \sqrt[3]{2} + \sqrt[3]{4})S^{(4)}p^3\right]\delta\tau^5 + O(\delta\tau^6) \end{aligned}$$

and the second-order PQP Campostrini wiggle gives

$$\begin{aligned} q' = & q + p\delta\tau - \frac{1}{2}S'\delta\tau^2 - \frac{1}{6}S''p\delta\tau^3 + \frac{1}{24}[S''S' - S'''p^2]\delta\tau^4 \\ & + \frac{1}{120}\left[\left(3S'''S' + S''^2\right)p - S^{(4)}p^3\right]\delta\tau^5 \\ & + \frac{1}{720}\left[-(3S'''S' + S''^2)S' + \left(5S'''S'' + 6S^{(4)}S'\right)p^2 - S^{(5)}p^4\right]\delta\tau^6 + O(\delta\tau^7) \end{aligned}$$

which agrees with the exact evolution operator $e^{\delta\tau \hat{H}}$ up to $O(\delta\tau^7)$.

IV. EQUILIBRIUM DISTRIBUTIONS

We now want to address the question as to what fixed point (equilibrium) distribution is produced by the Langevin [33, 34, 35] and Hybrid algorithms. The existence of such a fixed point distribution and its uniqueness follow from their ergodicity. We are, of course, looking for a fixed point distribution of q alone, and not of both q and p .

Since in general the momentum dependence of H' is not exactly Gaussian we do not expect either e^{-H} or $e^{-H'}$ to be an exact fixed point of a Hybrid Markov process. As we expect the equilibrium distribution to be close to the desired one, e^{-S} , it seems reasonable to parameterize it as $e^{-(S+\Delta S)}$. The condition that this is a fixed point is

$$\begin{aligned} & e^{-(S(q')+\Delta S(q'))} \\ &= \int dq e^{-(S(q)+\Delta S(q))} \int dp e^{-\frac{1}{2}p^2} \delta(q' - q'') \\ &= \int dq dp e^{-H(q,p)-\Delta S(q)} \delta(q' - q''), \end{aligned}$$

where $\begin{pmatrix} q'' \\ p'' \end{pmatrix} = U(\tau) \begin{pmatrix} q \\ p \end{pmatrix}$ is the phase space point reached at the end of a trajectory of length τ .

To solve this equation we change variables to (q'', p'') , whence we obtain

$$\begin{aligned} & e^{-(S(q')+\Delta S(q'))} \\ &= \int dq'' dp'' (\det U_*)^{-1} e^{-(H+\Delta S)\circ U^{-1}} \delta(q' - q''). \end{aligned}$$

If U is area-preserving then its Jacobian $\det U_* \equiv \det \frac{\partial(q'', p'')}{\partial(q, p)} \equiv e^{\text{tr} \ln U_*} = 1$, and if it is reversible then $U^{-1} = F \circ U \circ F$ where F is the momentum-flip operation $F : p \mapsto -p$.

In the general case we introduce the operation $\bar{\delta} : \Omega \mapsto \Omega \circ U^{-1} - \Omega \circ F \circ U \circ F$, which is zero if U is reversible, so

$$(H + \Delta S) \circ U^{-1} = (H + \Delta S) \circ F \circ U \circ F + \bar{\delta}(H + \Delta S).$$

Next we note that H is an even function of p , namely $H \circ F = H$, and ΔS does not depend on p , so $\Delta S \circ F = \Delta S$ also, hence

$$(H + \Delta S) \circ U^{-1} = (H + \Delta S) \circ U \circ F + \bar{\delta}(H + \Delta S).$$

In terms of the operator $\delta : \Omega \mapsto \Omega \circ U \circ F - \Omega \circ F$, which measures the change in some quantity Ω over a trajectory,⁷

$$\begin{aligned} (H + \Delta S) \circ U^{-1} &= (H + \Delta S) \circ F + \delta(H + \Delta S) + \bar{\delta}(H + \Delta S) \\ &= H + \Delta S + (\delta + \bar{\delta})(H + \Delta S). \end{aligned}$$

Therefore

$$\begin{aligned} &e^{-\left(S(q') + \Delta S(q')\right)} \\ &= \int dq'' dp'' e^{-(H + \Delta S)} e^{-(\delta + \bar{\delta})(H + \Delta S) - \text{tr} \ln U_*} \delta(q' - q'') \\ &= e^{-\left(S(q') + \Delta S(q')\right)} \int dp'' e^{-\frac{1}{2}p''^2} e^{-(\delta + \bar{\delta})(H + \Delta S) - \text{tr} \ln U_*}, \end{aligned}$$

and thus we obtain the condition

$$\left\langle e^{-(\delta + \bar{\delta})(H + \Delta S) - \text{tr} \ln U_*} \right\rangle_p = 1. \quad (4)$$

Since H is extensive so is δH , and thus we can show order by order in $\delta\tau$ that ΔS is extensive too.

A. Langevin Algorithm

For a reversible and area-preserving integration scheme, such as a symmetric symplectic integrator, the equilibrium distribution must satisfy

$$\left\langle e^{-\delta(H + \Delta S)} \right\rangle_p = 1. \quad (5)$$

For the PQP Langevin integrator we have

$$\begin{aligned} \delta H &= \frac{1}{12} \{-p^3 S''' + 3p S' S''\} \delta\tau^3 + \frac{1}{24} \{-p^4 S^{(4)} \\ &\quad + 3p^2(2S' S''' + S''^2) - 3S'^2 S''\} \delta\tau^4 + O(\delta\tau^5), \end{aligned}$$

and

$$\delta\Delta S = p\Delta S' \delta\tau + \left\{ \frac{1}{2}p^2\Delta S'' - \frac{1}{2}S'\Delta S' \right\} \delta\tau^2 + O(\delta\tau^3). \quad (6)$$

If we expand the integrand of equation (5) we obtain to leading non-vanishing order in $\delta\tau$

$$\langle \delta H + \delta\Delta S \rangle_p \sim 0,$$

and thus

$$\begin{aligned} 0 &\sim -\frac{1}{2} \{S'\Delta S' - \Delta S''\} \delta\tau^2 \\ &\quad - \frac{1}{8} \{-S''^2 + S''S'^2 + S^{(4)} - 2S'''S'\} \delta\tau^4 \\ &\quad + O(\delta\tau^6). \end{aligned}$$

⁷ An extra momentum flip is included for convenience.

This differential equation has an integrating factor of e^{-S} ,

$$([\Delta S' + \frac{1}{4}(S''S' - S''')\delta\tau^2]e^{-S})' \sim 0,$$

so

$$\Delta S' + \frac{1}{4}(S''S' - S''')\delta\tau^2 \sim Ke^S$$

for some constant K . As we are only determining the asymptotic expansion of ΔS we set $K = 0$, whence we obtain

$$\Delta S' \sim \frac{1}{4}\{S''' - S'S''\}\delta\tau^2.$$

Up to an constant, which is fixed by the normalization of $e^{-(S+\Delta S)}$, this gives

$$\Delta S \sim \frac{1}{8}\{2S'' - S'^2\}\delta\tau^2 + O(\delta\tau^4)$$

for the equilibrium distribution corresponding to the PQP Langevin equation.

It is most important to realise that this is only an asymptotic expansion, and the exact solution for ΔS may also involve terms which are exponentially small in the integration step size $\delta\tau$. One way to understand this is to observe that equation (5) is a Gaussian integral over the momenta p , and that the domain of integration must include momenta $p \gg 1/\delta\tau$. In particular this means that the statement made in [3] that the shift in the equilibrium distribution corresponds only to the addition of irrelevant operators, and that therefore the $\delta\tau$ errors can be ignored if one computes quantities in the continuum limit, is erroneous because of the existence of these subleading relevant contributions.

If we write the asymptotic expansion as $\Delta S = \sum_{n \geq 2} \Delta S_n \delta\tau^n$ then the preceding calculation is easily extended to find that the next term satisfies the equation

$$\Delta S'_4 \sim \frac{1}{48} (S'''S'^2 + 2S''S''' - 2S^{(4)}S' + S^{(5)}). \quad (7)$$

While this does not give a closed-form expression for ΔS_4 for arbitrary S , it is obvious that if S is a polynomial in q then ΔS_4 and all the other ΔS_n are too. In general the fields q have multiple components, and equation (7) then becomes

$$(\Delta S_4)_i \sim \frac{1}{48} (S_{ijk}S_jS_k + 2S_{jk}S_{ijk} - 2S_{ijjk}S_k + S_{ijjkk})$$

with the obvious notation.

For the QPQ integrator the shift in the equilibrium distribution is

$$\begin{aligned} \Delta S_2 &= \frac{1}{8}(S'^2 - S''), \\ \Delta S'_4 &= \frac{1}{384} (24S''^2S' + 8S'''S'^2 - 20S''S''' \\ &\quad - 4S^{(4)}S' - S^{(5)}); \end{aligned}$$

whereas for the PQP Campostrini wiggle (q.v., Appendix C) the leading shift satisfies

$$\begin{aligned} \Delta S'_4 &= \frac{1}{288} \left(2(4 + 3\sqrt[3]{2} + 2\sqrt[3]{4})S''^2S' \right. \\ &\quad \left. - (6 + 4\sqrt[3]{2} + 5\sqrt[3]{4})S'''S'^2 - 6(4 + 3\sqrt[3]{2} + 2\sqrt[3]{4})S''S''' \right. \\ &\quad \left. + 2(6 + 5\sqrt[3]{2} + 5\sqrt[3]{4})S^{(4)}S' - (6 + 5\sqrt[3]{2} + 5\sqrt[3]{4})S^{(5)} \right). \end{aligned}$$

B. Hybrid Algorithm

For the Hybrid algorithm we cannot derive an explicit formula for the shifted equilibrium distribution in general, but we can easily see from equation (II A) that the leapfrog integrator conserves a Hamiltonian H' which differs from H by terms of $O(\delta\tau^2)$. This means that $\delta H = O(\delta\tau^2)$, and thus we deduce from equation (5) that $\delta\Delta S \sim O(\delta\tau^2)$. Unlike the Langevin case considered previously the change in ΔS over trajectory has no reason not to be of the same size as ΔS itself, and thus we find that $\Delta S \sim O(\delta\tau^2)$.

V. NOISY ALGORITHMS

One of the main advantages of the Langevin and Hybrid algorithms is that they can be used for some non-local actions where an unbiased stochastic estimate of the force⁸ Σ' can be computed relatively cheaply, with $\langle \Sigma'(q) \rangle_\eta = S'(q)$ where the average is over some “noise” η which is chosen independently for each step.

It is helpful to think of the dynamics as that of a system evolving in the presence of some fixed background noise field; we shall leave this dependence on η implicit except where necessary. The noisy force estimator corresponds to the discrete mapping e^{tP} of equation (2) becoming

$$e^{tP} : f(q, p) \mapsto f(q, p - t\Sigma'(q)),$$

and we thus find that the symmetric QPQ integrator conserves the Hamiltonian $H' \equiv H + \Delta H$ where⁹

$$\Delta H = \Sigma - S + \frac{1}{24} \left[-p^2 \Sigma'' + 2\Sigma'^2 \right] \delta\tau^2 + O(\delta\tau^4).$$

This means that the change in H over a trajectory $(q_i, p_i) \mapsto (q_f, p_f)$ is

$$\begin{aligned} \delta H &= H(q_f, p_f) - H(q_i, p_i) = [H(q_f, p_f) - H(q_i, p_i)] - [H'(q_f, p_f) - H'(q_i, p_i)] \\ &= [H(q_f, p_f) - H'(q_f, p_f)] - [H(q_i, p_i) - H'(q_i, p_i)] = -\Delta H(q_f, p_f) + \Delta H(q_i, p_i) \\ &= -\delta\Delta H = \delta S - \delta\Sigma - \frac{1}{24} \left[-p^2 \delta\Sigma'' + 2\delta(\Sigma')^2 \right] \delta\tau^2 + O(\delta\tau^4) \\ &= (S' - \Sigma')p\delta\tau + \frac{1}{2} \left[\Sigma'^2 - S'\Sigma' - p^2(\Sigma'' - S'') \right] \delta\tau^2 \\ &\quad + \frac{1}{24} \left[p^3(4S''' - 3\Sigma''') + 6p(2\Sigma''\Sigma - S'\Sigma'' - 2S''\Sigma') \right] \delta\tau^3 + O(\delta\tau^4) \end{aligned}$$

where the last line is for a single leapfrog step where we have used equation (3).

If we consider the quantity $e^{-\delta H}$ averaged over the noise η for a single leapfrog step we obtain

$$\begin{aligned} \langle e^{-\delta H} \rangle_\eta &= 1 - \langle \delta H \rangle_\eta + \frac{1}{2} \langle \delta H^2 \rangle_\eta + \dots \\ &= 1 - \frac{1}{2} \langle (\Sigma' - S')^2 \rangle_\eta (1 - p^2) \delta\tau^2 + O(\delta\tau^3). \end{aligned} \tag{8}$$

Observe that the coefficient of $(1 - p^2)\delta\tau^2$ is proportional to the variance of the noisy estimator Σ' , and thus can only vanish if the force is computed exactly. Thus, regardless of the order of the integrator used there will always be an $O(\delta\tau^2)$ leading error for a single MD step.

A. Noisy Langevin Algorithm

If we now use equation (5) to compute the equilibrium distribution for the noisy Langevin algorithm we must average equation (8) over a Gaussian distribution for p , thus we find that

$$\langle e^{-\delta H} \rangle_{\eta, p} = 1 + A \delta\tau^4 + O(\delta\tau^6),$$

where

$$\begin{aligned} A \equiv \frac{1}{16} \left[-8S' \langle \Sigma' \Sigma'' \rangle_\eta + 4 \langle \Sigma''^2 \rangle_\eta - 2S'' \langle \Sigma'^2 \rangle_\eta \right. \\ \left. + 8S'^2 S'' + 4 \langle \Sigma''' \Sigma' \rangle_\eta - S' S''' \right. \\ \left. - S^{(4)} - 2S'^4 - 2S''^2 + 2S'^2 \langle \Sigma'^2 \rangle_\eta \right]. \end{aligned}$$

⁸ We do not necessarily require a noisy estimator Σ for the action S itself to implement the equations of motion, but we will require one if we wish to construct an “exact” algorithm, q.v. §VF.

⁹ There are two contributions to ΔH : the first due to the use of an approximate integrator in the presence of the background noise field, the second due to the noise itself.

Since, for a noisy force equation (6) is now of the form

$$\delta\Delta S = p\Delta S' \delta\tau + \frac{1}{2}(p^2\Delta S'' - \Sigma'\Delta S')\delta\tau^2 + O(\delta\tau^3),$$

we immediately find that $\Delta S = O(\delta\tau^2)$, and more specifically it satisfies the equation $e^S (e^{-S}\Delta S')' = -2A\delta\tau^2$.

B. Noisy Hybrid Algorithm

In the case of the noisy Hybrid algorithm (for which there are $\tau/\delta\tau$ steps per momentum refreshment) we recall that the evolution is to be averaged over independent noise for each integration step. Since the momentum distribution is no longer Gaussian after taking a leapfrog step, the leading order term is not cancelled as it would be if we only did one MD step per trajectory. The fixed point condition now becomes

$$\begin{aligned} & \left\langle e^{-\delta(H+\Delta S)} \right\rangle_{p, \eta_1, \dots, \eta_N} \\ &= \left\langle \left\langle e^{-\delta H_1} \right\rangle_{\eta_1} \dots \left\langle e^{-\delta H_N} \right\rangle_{\eta_N} e^{-\delta\Delta S} \right\rangle_p = 1, \end{aligned}$$

where $\delta H = \delta H_1 + \dots + \delta H_N$. Thus the leading error of the noisy Hybrid algorithm is $\Delta S \sim O(\tau\delta\tau)$, as each step contributes an error of $O(\delta\tau^2)$ by equation (8). A way of reducing this error to $O(\tau\delta\tau^2)$ is the R algorithm [10], which we shall discuss in §V E.

Note that when we construct a Hybrid algorithm with a noisy force estimator there is no reason to expect there to be a nearby conserved “shadow” Hamiltonian, as averages of non-linear Poisson brackets will not be correct. Thus if the average error per step is $O(\delta\tau^2)$ we expect the errors at the end of a trajectory of $\tau/\delta\tau$ steps to be $O(\tau\delta\tau)$ and not just $O(\delta\tau)$.

C. Pseudofermions and the Φ Algorithm

We now turn our attention to the specific case of gauge theories in the presence of dynamical fermions. We start with the probability distribution for the gauge field U with the quadratic fermion contribution integrated out in favour of the fermionic determinant, $P(U) \propto e^{-S_G(U)} \det \mathcal{M}(U)$, where S_G is the pure gauge part of the action and \mathcal{M} is the fermion kernel. For the purpose of this discussion we ignore the pure gauge contribution to the action since this is a simple local quantity whose force can be computed exactly, and shall focus on the awkward fermion determinant. As usual we may replace the determinant with an integral over a complex pseudofermion [1] field Φ and write the joint probability distribution of the gauge and pseudofermion fields as $P(U, \Phi) \propto \exp \left\{ -[S_G(U) + \Phi^\dagger \mathcal{M}^{-1} \Phi] \right\} \equiv e^{-S_{\text{eff}}(U, \Phi)}$. We have taken the fermion kernel to be $\mathcal{M} \equiv M^\dagger M$ representing two flavours of fermion with Dirac operator M to allow for a simple implementation of the pseudofermion heatbath.

1. Φ Algorithm

The gauge field U corresponds to the variable q used in the general discussion before, and we will evolve it along a classical trajectory in the presence of a fixed pseudofermion background field Φ . We introduce a conjugate momentum field p in order to define a Hamiltonian $H(U, p)$, $P(U, p) \propto \exp \left\{ -\left[\frac{1}{2}p^2 + S_{\text{eff}}(U)\right] \right\} \equiv e^{-H(U, p)}$; the action $S_{\text{eff}}(U)$ takes the role of the potential in the Hamiltonian, and of course depends implicitly on the background pseudofermion field. Note that the pseudofermion force term is given by

$$\frac{\partial S_F}{\partial U} = \Phi^\dagger \frac{\partial \mathcal{M}^{-1}}{\partial U} \Phi = -\Phi^\dagger \mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \mathcal{M}^{-1} \Phi. \quad (9)$$

The Φ algorithm of Gottlieb *et al.* [10] is identical to the Hybrid algorithm described in §III with the addition of pseudofermion refreshment from a Gaussian heatbath before each MD trajectory. With regards to fermions, the Φ algorithm is restricted to describing N_f degenerate flavours of fermions, where N_f is the number of fermions described by the operator \mathcal{M} .

The Φ algorithm uses a QPQ integrator, as the evaluation of the pseudofermion force acting on the gauge fields required for the P step is only evaluated once in this case.

According to the general arguments presented in §IV the fixed point distribution of the Φ algorithm must satisfy equation (5). Performing an asymptotic expansion of this in powers of $\delta\tau$, and observing that $\delta H = O(\delta\tau^2)$ for any trajectory length τ , we deduce that $\delta\Delta S \sim O(\delta\tau^2)$. We thus have that $\Delta S \sim O(\delta\tau^2)$, so the Φ algorithm is accurate to this order.

2. χ Algorithm

A slight variation of the Φ algorithm is the χ algorithm. The difference is that in the latter the pseudofermion heatbath refreshment is performed before every MD step as opposed to before each MD trajectory in the former. We might expect the error per step to be $O(\delta\tau^2)$, leading to an error per trajectory of $O(\tau\delta\tau)$. However, the error per trajectory is in fact $O(\tau\delta\tau^2)$,¹⁰ a proof of this will follow from that of the R algorithm to be given in §V E.

We note in passing that the pseudofermion force for the χ algorithm is given by

$$\frac{\partial S_F}{\partial U} = - \left\langle \chi^\dagger \mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \mathcal{M}^{-1} \chi \right\rangle_\chi. \quad (10)$$

D. Non-local Actions and Noisy Hybrid Algorithms

There is considerable interest in having the number of fermion flavours unequal to that described by the fermion kernel (e.g., less than four flavours for staggered fermions). For such theories the required probability distribution is given by $P(U) \propto e^{-S_G(U)} \det \mathcal{M}(U)^n$, where the number of multiplets n determines the number of fermion flavours (e.g., $n = \frac{1}{2}N_f$ for Wilson fermions and $n = \frac{1}{4}N_f$ for staggered fermions). For $n \notin \mathbb{Z}$, the conventional pseudofermion approach fails because neither a non-integer power of the Dirac operator nor its derivative can be evaluated directly, which would be required to calculate the force.

1. R_0 Algorithm

An alternative to the pseudofermion approach is to rewrite the determinant in trace log form where the effective fermion action is $S_F = -n \text{tr} \ln \mathcal{M}$. The pseudofermionic force is replaced by a noisy estimator for the trace, since computing the trace exactly is prohibitively expensive. This force is written as

$$\begin{aligned} \frac{\partial S_F}{\partial U} &= -n \text{tr} \left[\frac{\partial \ln \mathcal{M}}{\partial U} \right] = -n \text{tr} \left[\mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \right] \\ &= -n \text{tr} \left[(M^\dagger M)^{-1} \frac{\partial \mathcal{M}}{\partial U} \right] \\ &= -n \text{tr} \left[M^{\dagger^{-1}} \frac{\partial \mathcal{M}}{\partial U} M^{-1} \right] \\ &= -n \left\langle \eta^\dagger M^{\dagger^{-1}} \frac{\partial \mathcal{M}}{\partial U} M^{-1} \eta \right\rangle_\eta, \end{aligned}$$

where η is a complex noise vector sampled from a Gaussian heatbath of unit variance. Defining an auxilliary field $\chi \equiv M^\dagger \eta$ the force becomes

$$\frac{\partial S_F}{\partial U} = -n \left\langle \chi^\dagger \mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \mathcal{M}^{-1} \chi \right\rangle_{\chi=M^\dagger \eta}. \quad (11)$$

With this formulation we can use the noisy Hybrid algorithm of §V to generate gauge field configurations corresponding to any number of flavours: the noisy estimator for the force Σ' being defined by $\frac{\partial S_F}{\partial U} \equiv -n \langle \Sigma' \rangle_\chi$. This is the R_0 algorithm of Gottlieb *et al.* [10] and has leading order error $O(\tau\delta\tau)$.

¹⁰ Similar to the Φ algorithm except that it grows with τ because there is no shadow Hamiltonian.

E. Reversibility, Area Preservation, and the R Algorithm

As can be seen from equations (9), (10), and (11), the Φ , χ , and R_0 algorithms have similar “pseudofermion” force terms despite their different derivations. Indeed, we introduced the χ algorithm to emphasise this similarity: in both the χ and R_0 algorithms the “pseudofermion” force is computed from Gaussian noise η for each MD step. The difference is that in the former the pseudofermion field $\chi = M^\dagger(U(t))\eta$ is calculated from a heatbath at the beginning of each MD step, whereas in the latter the auxilliary field $\chi = M^\dagger(U(t + \frac{1}{2}\delta\tau))\eta$ is calculated at the midpoint of each MD step (that is, at the same time as the force itself is evaluated for the integrator).

The χ algorithm has $O(\tau\delta\tau^2)$ errors for $n = 1$ multiplets, whereas the R_0 algorithm has $O(\tau\delta\tau)$ errors. However, for $n = 0$ multiplets (i.e., no fermions) the algorithms are identical and have errors of $O(\delta\tau^2)$. It seems reasonable to expect that the leading error has a linear dependence both on the time within the MD step at which the pseudofermions are generated from their heatbath and on the number of multiplets, so if we evaluate the pseudofermion field at time $t = \frac{1}{2}(1 - n)\delta\tau$ through the MD step for $0 \leq n \leq 1$ we should obtain an $O(\tau\delta\tau^2)$ algorithm. This is the R algorithm of Gottlieb *et al.* [10]. For two flavours of staggered fermions, this means evaluating the pseudofermion field a quarter way through each MD update. Note that this algorithm is neither reversible nor area-preserving.

To prove that the R algorithm does indeed have $O(\tau\delta\tau^2)$ leading order error we again look at the fixed point (equilibrium) distribution. The condition is that given in equation (4), and in this case we have neither area-preservation nor reversibility. We consider a single step of the R algorithm, where the auxilliary field χ is computed at a time $t = \frac{1}{2}(1 - \alpha)\delta\tau$ with α a free parameter. Expanding equation (4) to leading non-vanishing order we obtain

$$\langle(\delta + \bar{\delta})\Delta S\rangle_{p,\eta} \sim -\langle(\delta + \bar{\delta})H + \text{tr} \ln U_*\rangle_{p,\eta} + \dots$$

We can compute the leading contributions to this quantity as follows; the change in energy over a trajectory is

$$\langle\delta H\rangle_{p,\eta} = 2n(n + 2p^2\alpha) \text{tr} \left[\mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \right] \delta\tau^2,$$

where the $O(\delta\tau)$ term vanishes upon noise averaging. Taylor expanding the Jacobian for each update step gives

$$\langle\text{tr} \ln U_*\rangle_{p,\eta} = n\alpha \text{tr} \left[\mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \right] \delta\tau^2.$$

Finally, the leading order contribution to the quantity $\bar{\delta}U = U^{-1}(\delta\tau) - F \circ U \circ F$ that measures the lack of reversibility of the integrator is

$$\langle\bar{\delta}H\rangle_{p,\eta} = 2p^2n\alpha \text{tr} \left[\mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \right] \delta\tau^2.$$

We thus find that

$$\langle(\delta + \bar{\delta})\Delta S\rangle_{p,\eta} \sim -A\delta\tau^2 + O(\delta\tau^3)$$

where

$$A = \frac{1}{2}n(n - \alpha)(1 - p^2) \text{tr} \left[\mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \mathcal{M}^{-1} \frac{\partial \mathcal{M}}{\partial U} \right].$$

If we choose $\alpha = n$ the leading term cancels, and thus the leading error is $O(\tau\delta\tau^2)$ for an entire trajectory of $\tau/\delta\tau$ steps. Therefore, as claimed, the R algorithm has errors of $O(\tau\delta\tau^2)$, and thus so does the χ algorithm since it corresponds to the special case of the R algorithm with $n = 1$.

F. Exact Noisy Algorithms

It is interesting to consider whether the noisy algorithms described above can be made exact (in the sense of having no integrator step-size errors), and if so how.

With the addition of an accept/reject step after the MD step, the Φ algorithm becomes the exact Hybrid Monte Carlo (HMC) algorithm [7].

The noisy Hybrid algorithm can be made exact by including a noisy acceptance step [36, 37, 38, 39] after each MD integration step. Note that a trajectory is defined as being the MD evolution between momentum refreshments, and can consist of any number of MDMC steps, where an MDMC step is an MD step followed by a (noisy) acceptance test. Using just one acceptance test at the end of the molecular dynamics trajectory is not valid, since reversibility is violated because of the noise. Unfortunately this exact version of the algorithm suffers from a significantly longer autocorrelation time: this is because the momentum must be flipped if a rejection occurs at any of the $\tau/\delta\tau$ noisy acceptance tests required per trajectory, just as for the second-order Langevin (Kramer's) algorithm [27, 40, 41].

The χ algorithm can be made exact by the addition of a Metropolis acceptance step after each MD step, and including a momentum flip after every rejected MD update: however, the resulting algorithm suffers from the same problems as the exact noisy Hybrid algorithm discussed before.¹¹

The R_0 algorithm can be made exact through the addition of a noisy acceptance test, however, since the algorithm has scaling $O(\tau\delta\tau)$, this would require very small step sizes. Unfortunately the R algorithm cannot be made exact by adding a (noisy) Metropolis acceptance step, because the lack of area-preservation and reversibility preclude detail balance being satisfied.

Clearly, the use of exact algorithms is preferable to that of the inexact algorithms described in this work. For non-local actions (e.g., a non-integral number of fermion multiplets) the Rational Hybrid Monte Carlo algorithm [11, 12] is a good candidate: in this algorithm the fractional power of the fermion kernel that appears in the pseudofermion bilinear is replaced by a rational approximation that can be directly evaluated and differentiated. Hence a pseudofermion formulation can be used, and a Metropolis acceptance test can be added to render the algorithm exact.

VI. CONCLUSIONS

In this paper we have described how the use of a symplectic integrator for the numerical integration of Hamilton's equations of motion not only preserves the reversibility and area-preserving properties of their exact solution, but also conserves a "shadow" Hamiltonian close to the original one. We have shown how this conserved Hamiltonian may be computed as a power series in the integration step size $\delta\tau$ using the BCH formula for the Lie algebra of Poisson brackets.

We then considered Markov processes of the Hybrid type, in which molecular dynamics and momentum refreshment Markov steps are alternated. Since the fixed point of the molecular dynamics step does not coincide with that of the momentum heatbath neither can be the fixed point of the full Markov process. We have derived a general condition (equation (4)) for this fixed point distribution, which simplifies to equation (5) for the case of reversible and area preserving MD integrators, and shown how properties of this distribution can be found by expanding these conditions in powers of the integration step size $\delta\tau$. For the case of the Langevin algorithm, it was shown how the equilibrium distribution can be found explicitly to any order in $\delta\tau$, and why this is only an asymptotic expansion.

Finally we considered those algorithms which use a noisy estimate of the force. Here the leading order behaviour of these algorithms was found for the noisy Langevin and Hybrid algorithms. It was shown that in general noisy Hybrid algorithms have leading order error $O(\tau\delta\tau)$ regardless of the order of the numerical integrator, however, there are special cases where we can cancel this leading term through a judicious choice of when we evaluate the noise, i.e., the R algorithm. We also considered how to render these algorithms exact through the addition of a Metropolis acceptance test.

Acknowledgements

ADK would like to thank Stefan Sint, Ivan Horváth, Barry Trager, and José Figueroa-O'Farrill for helpful discussions. MAC is supported under NSF grant PHY-0427646.

¹¹ In fact, it is essentially the same algorithm.

APPENDIX A: HAMILTONIAN VECTOR FIELDS AND POISSON BRACKETS

We shall denote by Λ^k the set of antisymmetric multilinear k -forms that act on k -tuples of vectors in the tangent bundle $T\mathcal{M}$ over a manifold \mathcal{M} .

Definition 1 *The exterior derivative $d : \Lambda^k \rightarrow \Lambda^{k+1}$ is the unique linear transformation satisfying*

1. $df(v) = v(f)$ for any 0-form $f \in \Lambda^0$ and vector field $v \in T\mathcal{M}$,
2. $d^2 = 0$, and
3. $d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^{\deg \alpha} \alpha \wedge d\beta$ (the anti-Leibniz rule).

Lemma 1 *For any $\Omega \in \Lambda^k$*

$$\begin{aligned} d\Omega(v_0, \dots, v_k) &= \sum_i (-1)^i v_i \Omega(v_0, \dots, \hat{v}_i, \dots, v_k) \\ &+ \sum_{i < j} (-1)^{i+j} \Omega([v_i, v_j], v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_k), \end{aligned}$$

where \hat{v}_j indicates that the term v_j is omitted, and $[a, b] \in T\mathcal{M}$ is the commutator of two vector fields $a, b \in T\mathcal{M}$.

Proof. We may express Ω in a local coordinate patch as $\Omega = \frac{1}{k!} \Omega_{\mu_1, \dots, \mu_k} dq^{\mu_1} \wedge \dots \wedge dq^{\mu_k}$, and thus

$$\begin{aligned} d\Omega(v_0, \dots, v_k) &= \frac{1}{k!} \frac{\partial \Omega_{\mu_1, \dots, \mu_k}}{\partial q^{\mu_0}} dq^{\mu_0} \wedge dq^{\mu_1} \wedge \dots \wedge dq^{\mu_k}(v_0, \dots, v_k) \\ &= \sum_{i=0}^k (-1)^i \frac{\partial \Omega_{\mu_0, \dots, \hat{\mu}_i, \dots, \mu_k}}{\partial q^{\mu_i}} v_0^{\mu_0} v_1^{\mu_1} \dots \hat{v}_i^{\mu_i} \dots v_k^{\mu_k} \\ &= \sum_{i=0}^k (-1)^i v_i^{\mu_i} \frac{\partial}{\partial q^{\mu_i}} [\Omega_{\mu_0, \dots, \hat{\mu}_i, \dots, \mu_k} v_0^{\mu_0} \dots \hat{v}_i^{\mu_i} \dots v_k^{\mu_k}] \\ &\quad + \sum_{0 \leq i < j \leq k} (-1)^{i+j} \Omega_{\mu_0, \dots, \hat{\mu}_i, \dots, \mu_k} \left[v_i^{\mu_i} \frac{\partial v_j^{\mu_j}}{\partial q^{\mu_i}} - v_j^{\mu_j} \frac{\partial v_i^{\mu_i}}{\partial q^{\mu_j}} \right] v_0^{\mu_0} \dots \hat{v}_i^{\mu_i} \dots \hat{v}_j^{\mu_j} \dots v_k^{\mu_k} \\ &= \sum_i (-1)^i v_i \Omega(v_0, \dots, \hat{v}_i, \dots, v_k) + \sum_{i < j} (-1)^{i+j} \Omega([v_i, v_j], v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_k). \end{aligned}$$

The result is independent of the coordinate system used for this verification. ■

For any 1-form θ and 2-form ω this identity is

$$\begin{aligned} d\theta(a, b) &= a\theta(b) - b\theta(a) - \theta([a, b]); \\ d\omega(a, b, c) &= a\omega(b, c) + b\omega(c, a) + c\omega(a, b) \\ &\quad - \omega([a, b], c) - \omega([b, c], a) - \omega([c, a], b). \end{aligned} \tag{A1}$$

Definition 2 *The cotangent bundle $T^*\mathcal{M}$ has a symplectic structure if there is a non-singular closed fundamental 2-form ω .*

Definition 3 *For each 0-form F on $T^*\mathcal{M}$ there is a corresponding Hamiltonian vector field \hat{F} defined by $dF \equiv i_{\hat{F}}\omega$ where i is the interior produce; equivalently we may write this as $dF(x) = \omega(\hat{F}, x) \forall x \in T\mathcal{M}$.*

Definition 4 *The Poisson bracket of two 0-forms is*

$$\{A, B\} \equiv -\omega(\hat{A}, \hat{B}) \quad A, B \in \Lambda^0.$$

Lemma 2 *The action of a Hamiltonian vector field \hat{A} on a 0-form F is given by $\hat{A}F = \{A, F\}$.*

Proof. $\hat{A}F = dF(\hat{A}) = i_{\hat{F}}\omega(\hat{A}) = \omega(\hat{F}, \hat{A}) = \{A, F\}$. ■

Lemma 3 *The space of 0-forms on $T^*\mathcal{M}$ together with the Poisson bracket form a Lie algebra; that is the Poisson bracket satisfies*

- $\{A, A\} = 0$ and
- $\{A, \{B, C\}\} + \{B, \{C, A\}\} + \{C, \{A, B\}\} = 0$
(Jacobi identity).

Proof. The antisymmetry is obvious. To establish the Jacobi identity consider $d\omega(\hat{A}, \hat{B}, \hat{C})$ for three arbitrary Hamiltonian vector fields \hat{A} , \hat{B} and \hat{C} . Recalling that the fundamental 2-form is closed, $d\omega = 0$, and using equation (A1) we have

$$\begin{aligned} d\omega(\hat{A}, \hat{B}, \hat{C}) &= \hat{A}\omega(\hat{B}, \hat{C}) + \hat{B}\omega(\hat{C}, \hat{A}) + \hat{C}\omega(\hat{A}, \hat{B}) \\ &\quad + \omega([\hat{A}, \hat{B}], \hat{C}) + \omega([\hat{B}, \hat{C}], \hat{A}) \\ &\quad + \omega([\hat{C}, \hat{A}], \hat{B}) \\ &= 0. \end{aligned} \tag{A2}$$

Now, $\hat{A}\omega(\hat{B}, \hat{C}) = -\hat{A}\{B, C\}$ by the definition of the Poisson bracket, so $\hat{A}\omega(\hat{B}, \hat{C}) = -\{A, \{B, C\}\}$ by application of lemma 2.

Similarly,

$$\begin{aligned} \omega([\hat{A}, \hat{B}], \hat{C}) &= -\omega(\hat{C}, [\hat{A}, \hat{B}]) = -i_{\hat{C}}\omega([\hat{A}, \hat{B}]) \\ &= -dC([\hat{A}, \hat{B}]) = -[\hat{A}, \hat{B}]C \\ &= -(\hat{A}\hat{B} - \hat{B}\hat{A})C = -\hat{A}\{B, C\} + \hat{B}\{A, C\} \\ &= -\{A, \{B, C\}\} + \{B, \{A, C\}\}. \end{aligned}$$

The Jacobi identity follows upon substituting these relations into equation A2. ■

Theorem 1 *Let \hat{A} and \hat{B} be two Hamiltonian vector fields, then their commutator is also a Hamiltonian vector field and furthermore $[\hat{A}, \hat{B}] = \widehat{\{A, B\}}$.*

Proof. Consider the action of the commutator on an arbitrary $F \in \Lambda^0$, $[\hat{A}, \hat{B}]F = (\hat{A}\hat{B} - \hat{B}\hat{A})F = \{A, \{B, F\}\} - \{B, \{A, F\}\}$ upon applying lemma 2. Using the Jacobi identity we obtain $[\hat{A}, \hat{B}]F = -\{F, \{A, B\}\} = \{\{A, B\}, F\} = \widehat{\{A, B\}}F$. As this must hold $\forall F$ the result follows. ■

This argument may be carried out more explicitly in a local coordinate patch. Locally ω may always be written as $\sum_i dq_i \wedge dp_i$ (Darboux theorem) so

$$\begin{aligned} \omega(\hat{A}, x) &= dA(x) \\ \Rightarrow \left(\sum_i dq_i \wedge dp_i \right) (\hat{A}, x) &= \left(\frac{\partial A}{\partial q^i} dq_i + \frac{\partial A}{\partial p^i} dp_i \right) (x) \\ \Rightarrow \left(\sum_i dq_i \wedge dp_i \right) \left(a^j \frac{\partial}{\partial q^j} + \tilde{a}^j \frac{\partial}{\partial p^j}, x^k \frac{\partial}{\partial q^k} + \tilde{x}^k \frac{\partial}{\partial p^k} \right) \\ &= a^i \tilde{x}^i - \tilde{a}^i x^i = \frac{\partial A}{\partial q^i} x^i + \frac{\partial A}{\partial p^i} \tilde{x}^i; \end{aligned}$$

and as this must hold for arbitrary x we may identify

$$a^i = \frac{\partial A}{\partial p^i} \quad \text{and} \quad \tilde{a}^i = -\frac{\partial A}{\partial q^i},$$

and thus

$$\hat{A} = \sum_i \left(\frac{\partial A}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial A}{\partial q^i} \frac{\partial}{\partial p^i} \right).$$

The commutator of \hat{A} and \hat{B} is

$$\begin{aligned}
[\hat{A}, \hat{B}] &= \left[\sum_i \left(\frac{\partial A}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial A}{\partial q^i} \frac{\partial}{\partial p^i} \right), \sum_j \left(\frac{\partial B}{\partial p^j} \frac{\partial}{\partial q^j} - \frac{\partial B}{\partial q^j} \frac{\partial}{\partial p^j} \right) \right] \\
&= \sum_{i,j} \left[\left(\frac{\partial A}{\partial p^i} \frac{\partial^2 B}{\partial q^i \partial p^j} - \frac{\partial B}{\partial p^i} \frac{\partial^2 A}{\partial q^i \partial p^j} - \frac{\partial A}{\partial q^i} \frac{\partial^2 B}{\partial p^i \partial p^j} + \frac{\partial B}{\partial q^i} \frac{\partial^2 A}{\partial p^i \partial p^j} \right) \frac{\partial}{\partial q^j} \right. \\
&\quad \left. + \left(-\frac{\partial A}{\partial p^i} \frac{\partial^2 B}{\partial q^i \partial q^j} + \frac{\partial B}{\partial p^i} \frac{\partial^2 A}{\partial q^i \partial q^j} + \frac{\partial A}{\partial q^i} \frac{\partial^2 B}{\partial p^i \partial q^j} - \frac{\partial B}{\partial q^i} \frac{\partial^2 A}{\partial p^i \partial q^j} \right) \frac{\partial}{\partial p^j} \right] \\
&= \sum_j \left[\frac{\partial}{\partial p^j} \sum_i \left(\frac{\partial A}{\partial p^i} \frac{\partial B}{\partial q^i} - \frac{\partial A}{\partial q^i} \frac{\partial B}{\partial p^i} \right) \frac{\partial}{\partial q^j} - \frac{\partial}{\partial q^j} \left(\frac{\partial A}{\partial p^i} \frac{\partial B}{\partial q^i} - \frac{\partial A}{\partial q^i} \frac{\partial B}{\partial p^i} \right) \frac{\partial}{\partial p^j} \right] \\
&= \sum_j \left(\frac{\partial \{A, B\}}{\partial p^j} \frac{\partial}{\partial q^j} - \frac{\partial \{A, B\}}{\partial q^j} \frac{\partial}{\partial p^j} \right) = \widehat{\{A, B\}},
\end{aligned}$$

where the Poisson bracket is

$$\begin{aligned}
\{A, B\} &\equiv - \sum_i \left(\frac{\partial A}{\partial q^i} \frac{\partial B}{\partial p^i} - \frac{\partial A}{\partial p^i} \frac{\partial B}{\partial q^i} \right) \\
&= - \left(\sum_i dq_i \wedge dp_i \right) \left(\frac{\partial A}{\partial p^j} \frac{\partial}{\partial q^j} - \frac{\partial A}{\partial q^j} \frac{\partial}{\partial p^j}, \frac{\partial B}{\partial p^k} \frac{\partial}{\partial q^k} - \frac{\partial B}{\partial q^k} \frac{\partial}{\partial p^k} \right) = -\omega(\hat{A}, \hat{B}).
\end{aligned}$$

APPENDIX B: BASIC PROPERTIES OF LIE ALGEBRAS

Definition 5 Let K be a commutative ring with unit. A Lie Algebra over K is a K -module¹² L together with a K -bilinear mapping $L \times L \rightarrow L : (x, y) \mapsto [x, y]$ called a Lie bracket which satisfies

$$\begin{aligned}
[x, x] &= 0 \\
[x, [y, z]] + [y, [z, x]] + [z, [x, y]] &= 0 \quad (\text{Jacobi identity})
\end{aligned} \tag{B1}$$

for all $x, y, z \in L$.

Note that this implies that the Lie bracket is antisymmetric,

$$[x, y] + [y, x] = 0 \quad \forall x, y \in L$$

because $[x+y, x+y] = [x, x] + [x, y] + [y, x] + [y, y] = [x, y] + [y, x]$. The converse is also true unless K has characteristic 2.

If \mathcal{A} is an associative algebra over K , then it has a natural Lie algebra structure¹³ given by $[x, y] \equiv xy - yx$.

Lemma 4 For any given Lie algebra L there is a unique associative algebra \mathcal{A}_0 called the enveloping algebra of L with a Lie algebra homomorphism $\phi_0 : L \rightarrow \mathcal{A}_0$ defined by $\phi_0 : [x, y] \mapsto xy - yx$ which has the universal property that for all \mathcal{A} for which there is a Lie algebra homomorphism $\phi : L \rightarrow \mathcal{A}$ there is a unique algebra homomorphism $f : \mathcal{A}_0 \rightarrow \mathcal{A}$ such that $\phi = f \circ \phi_0$.

In other words the enveloping algebra “contains” all associative algebras which have L as their natural Lie algebra.

Proof. Let \mathcal{T} be the tensor algebra¹⁴ over L and \mathcal{I} be the ideal of \mathcal{T} generated by the elements $x \otimes y - y \otimes x - [x, y]$ ($\forall x, y \in L$); then $\mathcal{A}_0 = \mathcal{T}/\mathcal{I}$. ■

¹² Recall that if K is a field then a K -module is a linear space.

¹³ Note that the Jacobi identity holds automatically.

¹⁴ \mathcal{T} is the universal algebra over the module L . We shall denote multiplication in \mathcal{T} by \otimes .

1. Poincaré–Birkhoff–Witt Theorem

Not only is a Lie algebra contained in its enveloping algebra, but in fact the tensor algebra has a direct sum decomposition $\mathcal{T} = \mathcal{T}_s \oplus \mathcal{I}$, where \mathcal{T}_s is the subalgebra of \mathcal{T} consisting of all symmetric tensors, so $\mathcal{A}_0 \cong \mathcal{T}_s$. This is the content of the following

Theorem 2 (Poincaré–Birkhoff–Witt) *Let L be a Lie algebra over K which is a free K -module with a totally ordered basis (x_i) , and let \mathcal{A}_0 be its enveloping algebra. Then \mathcal{A}_0 is a free K -module with the set of ordered products $\phi_0(x_{i_1}) \dots \phi_0(x_{i_n})$ ($i_0 \leq \dots \leq i_n$) as a basis.*

Proof. Recall from the proof of Lemma 4 that $\mathcal{A}_0 = \mathcal{T}/\mathcal{I}$. If we define \mathcal{T}_s to be the submodule of \mathcal{T} spanned by the ordered products $x_{i_1} \otimes \dots \otimes x_{i_n}$ ($i_1 \leq \dots \leq i_n$), then what we must show is that \mathcal{T} is the direct sum of modules \mathcal{T}_s and \mathcal{I} , that is $\mathcal{T} = \mathcal{T}_s \oplus \mathcal{I}$ and $\mathcal{T}_s \cap \mathcal{I} = \emptyset$. We shall do this by showing that each element of \mathcal{T} has a unique decomposition into an element of \mathcal{T}_s and an element of the ideal \mathcal{I} . Clearly, for elements of \mathcal{T}_s or \mathcal{I} we must have

$$\begin{aligned} x_{i_1} \otimes \dots \otimes x_{i_n} &= \left\{ x_{i_1} \otimes \dots \otimes x_{i_n} \right\} \oplus 0 (i_1 \leq \dots \leq i_n), \\ \alpha \otimes (x \otimes y - y \otimes x - [x, y]) \otimes \beta \\ &= 0 \oplus \left\{ \alpha \otimes (x \otimes y - y \otimes x - [x, y]) \otimes \beta \right\} (\alpha, \beta \in \mathcal{T}). \end{aligned}$$

For any other basis element of \mathcal{T} where the factors are not in increasing order we have

$$\begin{aligned} \alpha \otimes y \otimes x \otimes \beta \\ &= \left\{ \alpha \otimes x \otimes y \otimes \beta - \alpha \otimes [x, y] \otimes \beta \right\} \\ &\quad \oplus \left\{ \alpha \otimes (y \otimes x - x \otimes y + [x, y]) \otimes \beta \right\}. \end{aligned}$$

where $x < y$. All that remains to show is that this decomposition is well-defined and does not depend upon the order in which we apply the preceding identity. The only non-trivial case occurs when there are three adjacent out-of-order factors, for which we have both

$$\begin{aligned} \alpha \otimes z \otimes y \otimes x \otimes \beta &= \left\{ \alpha \otimes z \otimes x \otimes y \otimes \beta - \alpha \otimes z \otimes [x, y] \otimes \beta \right\} \oplus \left\{ \alpha \otimes z \otimes (y \otimes x - x \otimes y + [x, y]) \otimes \beta \right\} \\ &= \left\{ \alpha \otimes x \otimes z \otimes y \otimes \beta - \alpha \otimes [x, z] \otimes y \otimes \beta - \alpha \otimes z \otimes [x, y] \otimes \beta \right\} \\ &\quad \oplus \left\{ \alpha \otimes (z \otimes x - x \otimes z + [x, z]) \otimes y \otimes \beta + \alpha \otimes z \otimes (y \otimes x - x \otimes y + [x, y]) \otimes \beta \right\} \\ &= \left\{ \alpha \otimes x \otimes y \otimes z \otimes \beta - \alpha \otimes x \otimes [y, z] \otimes \beta - \alpha \otimes [x, z] \otimes y \otimes \beta - \alpha \otimes z \otimes [x, y] \otimes \beta \right\} \\ &\quad \oplus \left\{ \alpha \otimes x \otimes (z \otimes y - y \otimes z + [y, z]) \otimes \beta + \alpha \otimes (z \otimes x - x \otimes z + [x, z]) \otimes y \otimes \beta \right. \\ &\quad \left. + \alpha \otimes z \otimes (y \otimes x - x \otimes y + [x, y]) \otimes \beta \right\} \end{aligned}$$

and

$$\begin{aligned} \alpha \otimes z \otimes y \otimes x \otimes \beta &= \left\{ \alpha \otimes y \otimes z \otimes x \otimes \beta - \alpha \otimes [y, z] \otimes x \otimes \beta \right\} \oplus \left\{ \alpha \otimes (z \otimes y - y \otimes z + [y, z]) \otimes x \otimes \beta \right\} \\ &= \left\{ \alpha \otimes y \otimes x \otimes z \otimes \beta - \alpha \otimes y \otimes [x, z] \otimes \beta - \alpha \otimes [y, z] \otimes x \otimes \beta \right\} \\ &\quad \oplus \left\{ \alpha \otimes y \otimes (z \otimes x - x \otimes z + [x, z]) \otimes \beta + \alpha \otimes (z \otimes y - y \otimes z + [y, z]) \otimes x \otimes \beta \right\} \\ &= \left\{ \alpha \otimes x \otimes y \otimes z \otimes \beta - \alpha \otimes [x, y] \otimes z \otimes \beta - \alpha \otimes y \otimes [x, z] \otimes \beta - \alpha \otimes [y, z] \otimes x \otimes \beta \right\} \\ &\quad \oplus \left\{ \alpha \otimes (y \otimes x - x \otimes y + [x, y]) \otimes z \otimes \beta + \alpha \otimes y \otimes (z \otimes x - x \otimes z + [x, z]) \otimes \beta \right. \\ &\quad \left. + \alpha \otimes (z \otimes y - y \otimes z + [y, z]) \otimes x \otimes \beta \right\}. \end{aligned}$$

These two values differ by

$$\begin{aligned}
& \left\{ \alpha \otimes x \otimes [y, z] \otimes \beta + \alpha \otimes [x, z] \otimes y \otimes \beta + \alpha \otimes z \otimes [x, y] \otimes \beta \right. \\
& \quad \left. - \alpha \otimes [x, y] \otimes z \otimes \beta - \alpha \otimes y \otimes [x, z] \otimes \beta - \alpha \otimes [y, z] \otimes x \otimes \beta \right\} \\
& \oplus \left\{ \alpha \otimes x \otimes [y, z] \otimes \beta + \alpha \otimes [x, z] \otimes y \otimes \beta + \alpha \otimes z \otimes [x, y] \otimes \beta \right. \\
& \quad \left. - \alpha \otimes [x, y] \otimes z \otimes \beta - \alpha \otimes y \otimes [x, z] \otimes \beta - \alpha \otimes [y, z] \otimes x \otimes \beta \right\} \\
& = \left\{ \alpha \otimes (x \otimes [y, z] - [y, z] \otimes x) \otimes \beta + \alpha \otimes ([x, z] \otimes y - y \otimes [x, z]) \otimes \beta + \alpha \otimes (z \otimes [x, y] - [x, y] \otimes z) \otimes \beta \right\} \\
& \quad \oplus \left\{ \alpha \otimes [x, z] \otimes y \otimes \beta + \alpha \otimes z \otimes [x, y] \otimes \beta - \alpha \otimes [x, y] \otimes z \otimes \beta - \alpha \otimes y \otimes [x, z] \otimes \beta - \alpha \otimes [y, z] \otimes x \otimes \beta \right\} \\
& = \left\{ \alpha \otimes ([x, [y, z]] + [y, [z, x]] + [z, [x, y]]) \otimes \beta \right\} \oplus \left\{ \alpha \otimes ([x, [y, z]] + [y, [z, x]] + [z, [x, y]]) \otimes \beta \right\} = 0
\end{aligned}$$

where we have used the Jacobi identity. ■

2. Free Lie Algebras

We are interested in the properties of the Lie algebra generated by some set of operators, but *a priori* we know nothing about the nature of the Lie brackets of these generators. We therefore wish to work in the context of the most general Lie algebra which can be constructed from these generators, for which all Lie brackets are assumed distinguishable unless they are related by the defining relations (B1): any further relations between Lie brackets may be applied *post facto*. More formally this means that we wish to carry out our calculations in the *free Lie algebra* over our set of generators.

Let L_0 be a Lie Algebra over K , and A a set with a mapping $i : A \rightarrow L_0$. L_0 is called *free* on A if for any Lie algebra L and any mapping $f : A \rightarrow L$ there is a unique Lie algebra homomorphism $\bar{f} : L_0 \rightarrow L$ such that $\bar{f} \circ i = f$. This is a universal definition, as shown by

Theorem 3 *For every set A there is a unique free Lie algebra $L(A)$ on A , $L(A)$ is a naturally graded K -module, i is an injection, the component of $L(A)$ of degree 1 is the free submodule generated by $i(A)$, and $L(A)$ is generated as a Lie algebra by A .*

3. Hall Bases

Central to calculations in free Lie algebras is the question of how many independent basis elements are there of a given degree, and how to reduce an arbitrary expression to canonical form in terms of such a basis.

Definition 6 (Hall Trees [42, 43, 44, 45]) *Given a set A consider the set of all binary trees with leaf nodes labelled by elements of A : this set is called the free Magma $M(A)$. We shall denote the tree h whose left subtree is h' and whose right subtree is h'' by $h = (h', h'')$. A Hall set H is a totally ordered subset of $M(A)$ containing A which satisfies*

$$\begin{aligned}
& h < h'' \quad \forall h = (h', h'') \in H - A \\
& \quad \quad \quad h = (h', h'') \in H \\
& \iff \begin{cases} h', h'' \in H \text{ and } h' < h'' \text{ and} \\ \text{either } h' \in A \text{ or } h' = (x, y) \text{ and } y \geq h''. \end{cases}
\end{aligned}$$

There is map $f : M(A) \rightarrow L(A)$ defined by $f(a) \mapsto a$ if $a \in A$, and $f : (h', h'') \mapsto [f(h'), f(h'')]$; the result of applying this map to a Hall tree gives a *Hall word*, and the Hall words corresponding to any Hall set form a *Hall basis* for $L(A)$ (as a K -module).

We shall use the following Hall basis for our calculations: $x \in H$ iff

$$\begin{aligned}
& x \in A \\
& \text{or } x = [y, z] \quad y, z \in H, y < z \\
& \text{or } x = [y, [z, u]] \quad y, z, u \in H, z < y \\
& \text{and } x < y \quad \text{if } \deg x < \deg y.
\end{aligned}$$

Any expression built out of Lie brackets may be reduced to canonical form by the applying the following transformations:

$$\begin{aligned} [sx + ty, z] &\mapsto s[x, y] + t[y, z] && \text{where } s, t \in K \\ [z, y] &\mapsto -[y, z] && \text{if } y < z \\ [y, [z, u]] &\mapsto -[u, [y, z]] + [z, [y, u]] && \text{if } y < z < u, \end{aligned}$$

with the elements of A themselves ordered lexicographically (i.e., alphabetically).

For generating set $\{A, B\}$ this leads to the following basis:

$$\begin{aligned} &\{A, B\}, \quad \{[A, B]\}, \\ &\{[A, [A, B]], [B, [A, B]]\}, \\ &\{[A, [A, [A, B]]], [B, [A, [A, B]]], [B, [B, [A, B]]]\}, \\ &\{[[A, B], [B, [A, B]]], [B, [B, [B, [A, B]]]], \\ &\quad [A, [A, [A, [A, B]]]], [B, [A, [A, [A, B]]]], \\ &\quad [[A, B], [A, [A, B]]], [B, [B, [A, [A, B]]]]\}, \\ &\dots \end{aligned}$$

4. Dimension of Hall Bases

The dimension of the Hall basis of degree N on a set of cardinality q is given by Witt's formula [46] $a_N = \frac{1}{N} \sum_{d|N} \mu(d) q^{N/d}$, where μ is the Möbius function.

By the Poincaré–Birkhoff–Witt theorem the set of ordered (symmetric) monomials on the independent commutators is a basis for the universal enveloping algebra of the free Lie algebra. A basis for words of length N is therefore provided by symmetric products of n_k words of length k chosen from the a_k generators of the free Lie algebra, where $\sum_{k \geq 1} k n_k = N$. There are exactly $(-)^{n_k} \binom{-a_k}{n_k} = \binom{a_k + n_k - 1}{n_k}$ ways of choosing these n_k words symmetrically, and this is the coefficient of x^k in the series expansion of $(1 - x^k)^{-a_k} = \sum_{n_k \geq 0} \binom{-a_k}{n_k} (-x^k)^{n_k}$. The total number of words of length N is thus the coefficient of x^N in the generating function $g \equiv \prod_{k \geq 1} (1 - x^k)^{-a_k}$. On the other hand, the universal enveloping algebra is just the free algebra on q symbols, so there are q^N independent basis elements for words of length N , and thus a_k is determined from the equation $g = \sum_{k=0}^{\infty} q^k x^k = (1 - qx)^{-1}$. Witt's solution is obtained by taking the logarithm of this equation, $-\sum_{k \geq 1} a_k \ln(1 - x^k) = -\ln(1 - qx)$, and equating the coefficients of x^N , $\sum_{d|N} a_d / (N/d) = q^N / N$. Using the Möbius inversion formula we obtain $a_N = \frac{1}{N} \sum_{d|N} \mu(d) q^{N/d}$.

The number of independent commutators on q letters is therefore

$$\begin{aligned} a_1^{(q)} &= q, \\ a_2^{(q)} &= \frac{1}{2}q(q-1), \\ a_3^{(q)} &= \frac{1}{3}q(q-1)(q+1), \\ a_4^{(q)} &= \frac{1}{4}q^2(q-1)(q+1), \\ a_5^{(q)} &= \frac{1}{5}q(q-1)(q+1)(q^2+1), \\ a_6^{(q)} &= \frac{1}{6}q(q-1)(q+1)(q^3+q-1), \\ &\dots \end{aligned}$$

More specifically we have $a_1^{(2)} = 2$, $a_2^{(2)} = 1$, $a_3^{(2)} = 2$, $a_4^{(2)} = 3$, $a_5^{(2)} = 6$, $a_6^{(2)} = 9$, $a_7^{(2)} = 18$, $a_8^{(2)} = 30$, $a_9^{(2)} = 56$, $a_{10}^{(2)} = 99$, $a_{11}^{(2)} = 186$, $a_{12}^{(2)} = 335$, and so on.

5. Baker–Campbell–Hausdorff Formula

Theorem 4 ([42]) *Let M be an associative algebra over a commutative ring $K \supset \mathbb{Q}$ and let d be a derivation on M . For any power series $f(X) = \sum_{n \geq 0} a_n X^n$ we have*

$$df(X) = \sum_{k \geq 1} \frac{1}{k!} (\text{ad } X)^{k-1} (dX) f^{(k)}(X). \quad (\text{B2})$$

where $\text{ad } a : b \mapsto [a, b]$.

Proof. Equation (B2) is linear in f , so it suffices to consider $f(X) = X^n$, for which we have

$$dX^n = \sum_{k=1}^n \binom{n}{k} (\text{ad } X)^{k-1} (dX) X^{n-k}.$$

We shall prove this by induction on n : for $n = 0$ it is trivially true, and

$$dX^{n+1} = d(X X^n) = dX X^n + X dX^n = dX X^n + X \sum_{k=1}^n \binom{n}{k} (\text{ad } X)^{k-1} (dX) X^{n-k}.$$

Using the identity $Xu = [X, u] + uX = (\text{ad } X)u + uX$ we obtain

$$\begin{aligned} dX^{n+1} &= dX X^n + \sum_{k=1}^n \binom{n}{k} (\text{ad } X)^k (dX) X^{n-k} + \sum_{k=1}^n \binom{n}{k} (\text{ad } X)^{k-1} (dX) X^{n+1-k} \\ &= dX X^n + \sum_{k=2}^{n+1} \binom{n}{k-1} (\text{ad } X)^{k-1} (dX) X^{n+1-k} + \sum_{k=1}^n \binom{n}{k} (\text{ad } X)^{k-1} (dX) X^{n+1-k} \\ &= dX X^n + \sum_{k=2}^n \left[\binom{n}{k-1} + \binom{n}{k} \right] (\text{ad } X)^{k-1} (dX) X^{n+1-k} + (\text{ad } X)^n (dX) + n dX X^n \\ &= \sum_{k=1}^{n+1} \binom{n+1}{k} (\text{ad } X)^{k-1} (dX) X^{n+1-k}. \end{aligned}$$

■

Corollary 1

$$de^X = g(\text{ad } X)(dX)e^X$$

where

$$g(x) = \sum_{k \geq 1} \frac{x^{k-1}}{k!} = \frac{e^x - 1}{x}.$$

Proof. By Theorem 4 we have

$$de^X = \sum_{k \geq 1} \frac{1}{k!} (\text{ad } X)^{k-1} (dX) e^X = g(\text{ad } X)(dX) e^X.$$

■

Definition 7 *We define the Hausdorff series as*

$$H \equiv \sum_{n \geq 1} c_n(A, B) \quad \text{where} \quad e^H \equiv e^A e^B,$$

and the c_n are homogeneous of degree n in the generators A and B .

Theorem 5 Using the derivations¹⁵ d_A and d_B defined by $d_AA = A$, $d_AB = 0$, $d_BA = 0$, and $d_BB = B$, and setting $d \equiv d_A + d_B$, we have

$$dH = \left(\frac{\text{ad } H}{2} \coth \frac{\text{ad } H}{2} \right) (A + B) - \left(\frac{\text{ad } H}{2} \right) (A - B).$$

For any derivation D we have $D1 = D(1 \cdot 1) = D1 \cdot 1 + 1 \cdot D1 = 2D1$, so $D1 = 0$. It is also trivial to verify inductively that $d_X X^n = nX^n$ and $d_X e^X = X e^X$.

Proof. Using Corollary 1 we obtain

$$\begin{aligned} d_A e^H &= d_A(e^A e^B) = d_A e^A e^B + e^A d_A e^B \\ &= A e^H = g(\text{ad } H)(d_A H) e^H, \\ d_B e^{-H} &= d_B(e^{-B} e^{-A}) = d_B e^{-B} e^{-A} + e^{-B} d_B e^{-A} \\ &= -B e^{-H} = g(\text{ad}(-H))(d_B(-H)) e^{-H}; \end{aligned}$$

whence

$$\begin{aligned} A &= g(\text{ad } H)(d_A H), \\ B &= g(-\text{ad } H)(d_B H) \end{aligned}$$

and

$$\begin{aligned} d_A H &= [g(\text{ad } H)]^{-1} A, \\ d_B H &= [g(-\text{ad } H)]^{-1} B. \end{aligned}$$

We now observe that $1/g(x)$ may be decomposed into the sum of an even and an odd function

$$\frac{1}{g(\pm x)} = \frac{x}{2} \left(\coth \frac{x}{2} \mp 1 \right),$$

and the desired result follows immediately since $d = d_A + d_B$. ■

Definition 8 The Bernoulli numbers B_n are defined by

$$\frac{1}{g(x)} = \frac{x}{e^x - 1} \equiv \sum_{n \geq 0} \frac{B_n x^n}{n!}.$$

Taking the symmetric part of this expression we see that

$$\frac{x}{2} \coth \frac{x}{2} = \sum_{m \geq 0} \frac{B_{2m} x^{2m}}{(2m)!}.$$

Theorem 6 ([47, 48, 49, 50, 51, 52]) The terms in the Hausdorff series $c_n(A, B)$ are given by the recursion relations

$$c_{n+1} = \frac{1}{n+1} \left\{ -\frac{1}{2} [c_n, A - B] + \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{B_{2m}}{(2m)!} \sum_{\substack{k_1, \dots, k_{2m} \geq 1 \\ k_1 + \dots + k_{2m} = n}} [c_{k_1}, [\dots, [c_{k_{2m}}, A + B] \dots]] \right\}.$$

Proof. Using the properties that the c_n are homogeneous in A and B and that $\text{ad } 1 = 0$ we have

$$\begin{aligned} dH &= \sum_{n \geq 0} d c_n = \sum_{n \geq 1} n c_n, \\ \text{ad } H &= \sum_{k \geq 1} \text{ad } c_k; \end{aligned}$$

¹⁵ If $K = \mathbb{R}$ these definitions are equivalent to $d_A f(A, B) = \frac{\partial f(tA, B)}{\partial t} \Big|_{t=1}$, $d_B f(A, B) = \frac{\partial f(A, tB)}{\partial t} \Big|_{t=1}$ and $df(A, B) = \frac{\partial f(tA, tB)}{\partial t} \Big|_{t=1}$.

hence by Theorem 5

$$\begin{aligned}
dH &= \left(\frac{\text{ad } H}{2} \coth \frac{\text{ad } H}{2} \right) (A + B) - \left(\frac{\text{ad } H}{2} \right) (A - B) \\
&= \sum_{m \geq 0} \frac{B_{2m}}{(2m)!} (\text{ad } H)^{2m} (A + B) - \left(\frac{\text{ad } H}{2} \right) (A - B) \\
&= \sum_{m \geq 0} \frac{B_{2m}}{(2m)!} \left(\sum_{k \geq 1} \text{ad } c_k \right)^{2m} (A + B) - \left(\frac{\text{ad } H}{2} \right) (A - B)
\end{aligned}$$

and so

$$\sum_{n \geq 0} (n+1) c_{n+1} = \sum_{n \geq 1} \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{B_{2m}}{(2m)!} \sum_{\substack{k_1, \dots, k_{2m} \geq 1 \\ k_1 + \dots + k_{2m} = n}} \text{ad } c_{k_1} \dots \text{ad } c_{k_{2m}} (A + B) - \sum_{n \geq 1} \frac{1}{2} (\text{ad } c_n) (A - B)$$

therefore, equating terms of equal gradation,

$$(n+1) c_{n+1} = \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{B_{2m}}{(2m)!} \sum_{\substack{k_1, \dots, k_{2m} \geq 1 \\ k_1 + \dots + k_{2m} = n}} \text{ad } c_{k_1} \dots \text{ad } c_{k_{2m}} (A + B) - \frac{1}{2} (\text{ad } c_n) (A - B).$$

■

The first few terms in the Hausdorff series are

$$\begin{aligned}
\ln(e^A e^B) &= \{A + B\} + \frac{1}{2} [A, B] + \frac{1}{12} \left\{ [A, [A, B]] - [B, [A, B]] \right\} \\
&\quad - \frac{1}{24} [B, [A, [A, B]]] \\
&\quad + \frac{1}{720} \left\{ \begin{aligned} &-4 [B, [A, [A, [A, B]]]] - 6 [[A, B], [A, [A, B]]] \\ &+ 4 [B, [B, [A, [A, B]]]] - 2 [[A, B], [B, [A, B]]] \\ &- [A, [A, [A, [A, B]]]] + [B, [B, [B, [A, B]]]] \end{aligned} \right\} + \dots
\end{aligned} \tag{B3}$$

From this we easily obtain the formula for a symmetric product

$$\begin{aligned}
\ln(e^{A/2} e^B e^{A/2}) &= \{A + B\} - \frac{1}{24} \left\{ 2[B, [A, B]] + [A, [A, B]] \right\} \\
&\quad + \frac{1}{5760} \left\{ \begin{aligned} &32 [B, [B, [A, [A, B]]]] - 16 [[A, B], [B, [A, B]]] \\ &+ 28 [B, [A, [A, [A, B]]]] + 12 [[A, B], [A, [A, B]]] \\ &+ 8 [B, [B, [B, [A, B]]]] + 7 [A, [A, [A, [A, B]]]] \end{aligned} \right\} + \dots
\end{aligned} \tag{B4}$$

Since $e^A e^B = e^{A+B+\delta(A,B)}$ and $e^{-B} e^{-A} = e^{-B-A+\delta(-B,-A)}$ we see that $\delta(A, B) = -\delta(-B, -A)$, so under interchange of A and B all terms of even grading in change sign, whereas those of odd grading do not.

APPENDIX C: HIGHER-ORDER SYMMETRIC SYMPLECTIC INTEGRATORS

It was observed by Campostrini [13, 53] that one can construct higher-order integrators by the following method [14]: since $U_0(\delta\tau) = e^{\delta\tau\hat{H}} + R_0\delta\tau^3 + O(\delta\tau^5)$ we observe that the “wiggles”

$$U_0(\epsilon)U_0(-\sigma\epsilon)U_0(\epsilon) = e^{\epsilon(2-\sigma)\hat{H}} + R_0(2-\sigma^3)\epsilon^3 + O(\delta\tau^5)$$

can be adjusted to give an integration scheme correct to $O(\delta\tau^5)$ by choosing $\sigma = \sqrt[3]{2}$. The step size may be kept fixed by taking $\epsilon = \delta\tau/(2-\sigma)$.

Naturally, this wiggle may itself be iterated to give integration schemes of arbitrarily high order, all of which are still reversible and area-preserving. We use the recursive definition

$$\begin{aligned}
&U_n(\epsilon_n)U_n(-\sigma_n\epsilon_n)U_n(\epsilon_n) \\
&= e^{\epsilon(2-\sigma_n)\hat{H}} + R_n(2-\sigma_n^{2n+1})\epsilon_n^{2n+1} + O(\delta\tau^{2n+3})
\end{aligned}$$

and choose $\sigma_n = \sqrt[n+1]{2}$ and $\epsilon_n = \delta\tau/(2 - \sigma_n)$.

It was noted by Yoshida [16] that while the lowest-order Campostrini scheme is optimal, in the sense of requiring the fewest integration steps, the second order Campostrini wiggle is not. The lowest order Campostrini scheme corresponds to the operator

$$\begin{aligned} & e^{\epsilon Q/2} e^{\epsilon P} e^{\epsilon Q/2} e^{-\sigma\epsilon Q/2} e^{-\sigma\epsilon P} e^{-\sigma\epsilon Q/2} e^{\epsilon Q/2} e^{\epsilon P} e^{\epsilon Q/2} \\ & = e^{\epsilon Q/2} e^{\epsilon P} e^{(1-\sigma)\epsilon Q/2} e^{-\sigma\epsilon P} e^{(1-\sigma)\epsilon Q/2} e^{\epsilon P} e^{\epsilon Q/2} \end{aligned}$$

which uses precisely seven steps. The second order wiggle uses 19 steps, while the 15 step operator

$$\begin{aligned} & U_0(u \delta\tau) U_0(v \delta\tau) U_0(w \delta\tau) U_0(x \delta\tau) \\ & U_0(w \delta\tau) U_0(v \delta\tau) U_0(u \delta\tau) \\ & = e^{u \delta\tau Q/2} e^{u \delta\tau P} e^{(u+v) \delta\tau Q/2} e^{v \delta\tau P} e^{(v+w) \delta\tau Q/2} \\ & \quad e^{w \delta\tau P} e^{(w+x) \delta\tau Q/2} e^{x \delta\tau P} e^{(w+x) \delta\tau Q/2} \\ & \quad e^{w \delta\tau P} e^{(v+w) \delta\tau Q/2} e^{v \delta\tau P} e^{(u+v) \delta\tau Q/2} \\ & \quad e^{u \delta\tau P} e^{u \delta\tau Q/2} \end{aligned}$$

has errors of $O(\delta\tau^7)$ if the ideal defined by

$$\begin{aligned} x^5 + 2w^5 + 2v^5 + 2u^5 &= 0, \\ x^3 + 2w^3 + 2v^3 + 2u^3 &= 0, \\ x + 2w + 2v + 2u &= 1, \\ 4xwv^3 - 2xw^3v - 2x^3vu + 4xwu^3 - 4w^3vu \\ - 2x^3wu - 2xw^3u - 2x^3wv + 8wvu^3 - 2xv^3u \\ + 4xvu^3 - x^4w + x^2w^3 - 4wv^3u - 2w^3v^2 \\ - 2v^3u^2 - 2w^3u^2 + 2wu^4 + 2vu^4 + 4v^2u^3 \\ - 4w^4u + 4w^2u^3 - 4v^4u + 2wv^4 - x^3u^2 \\ + xu^4 - x^4u + x^2u^3 - 4w^4v + 4w^2v^3 \\ + x^2v^3 - x^4v + xv^4 - x^3v^2 - x^3w^2 + xw^4 &= 0 \end{aligned}$$

is not empty. This is indeed the case, as a Gröbner basis computation shows that v , w , and x may be expressed as polynomials in u , which is a root of the following irreducible polynomial

$$\begin{aligned} & 5632424294400000000u^{39} - 92200336819200000000u^{38} \\ & + 7106323614105600000000u^{37} - 3437629764814080000000u^{36} \\ & + 117450379289433600000000u^{35} - 302602294524211200000000u^{34} \\ & + 613444683393285120000000u^{33} - 100894346480650176000000u^{32} \\ & + 137871545973425856000000u^{31} - 159597428255349696000000u^{30} \\ & + 159057766014056179200000u^{29} - 138323253491741289600000u^{28} \\ & + 106099417410611328000000u^{27} - 72361810116050054400000u^{26} \\ & + 44125123305044761920000u^{25} - 24138408506765309280000u^{24} \\ & + 11867110408796028480000u^{23} - 5247011965321527840000u^{22} \\ & + 2086800152523757920000u^{21} - 746466059135744064000u^{20} \\ & + 240110266627607904000u^{19} - 69431877142547472000u^{18} \\ & + 18041618760883056000u^{17} - 4210061488653312000u^{16} \\ & + 881426634100156800u^{15} - 165335574305894400u^{14} \\ & + 27731884779770400u^{13} - 4148250096765600u^{12} \\ & + 551410054740000u^{11} - 64829840769360u^{10} \\ & + 6700635295200u^9 - 604016460000u^8 \\ & + 46995575760u^7 - 3112757280u^6 \\ & + 172255032u^5 - 7756920u^4 \\ & + 273360u^3 - 7080u^2 \\ & + 120u - 1. \end{aligned}$$

This polynomial has three real roots, corresponding to the numerical solutions found by Yoshida:

$$\begin{aligned}
 u &= 0.784513610477557263819497633866, & v &= 0.235573213359358133684793198233, \\
 w &= -1.177679984178871006946415596562, & x &= 1.315186320683911218884249687935; \\
 u &= 1.439848167976783090930499281479, & v &= 0.004260681870792016799146793392, \\
 w &= -2.132285222001451515523597811357, & x &= 2.376352744307752823717294656042; \\
 u &= 1.447782562399297932897896663298, & v &= -2.144035316305389310213622103136, \\
 w &= 0.001528862284249274922787428878, & x &= 2.389447783243684212186399178641.
 \end{aligned}$$

-
- [1] D. H. Weingarten and D. N. Petcher, Phys. Lett. **B99**, 333 (1981).
 - [2] F. Fucito, E. Marinari, G. Parisi, and C. Rebbi, Nucl. Phys. **B180** [FS2], 369 (1981).
 - [3] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. **D32**, 2736 (1985).
 - [4] D. J. E. Callaway and A. Rahman, Phys. Rev. Lett. **49**, 613 (1982).
 - [5] J. Polonyi and H. W. Wyld, Phys. Rev. Lett. **51**, 2257 (1983), erratum: ibid. 52:401, 1984.
 - [6] S. Duane, Nucl. Phys. **B257** [FS14], 652 (1985).
 - [7] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. **195B**, 216 (1987).
 - [8] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
 - [9] P. J. Rossky, J. D. Doll, and H. L. Friedman, J. Chem. Phys. **69**, 4628 (1978).
 - [10] S. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Phys. Rev. **D35**, 2531 (1987).
 - [11] M. A. Clark and A. D. Kennedy, Phys. Rev. Lett. **98**, 051601 (2007), hep-lat/0608015.
 - [12] M. A. Clark and A. D. Kennedy, Phys. Rev. **D75**, 011502 (2007), hep-lat/0610047.
 - [13] M. Campostrini and P. Rossi, Nucl. Phys. **B329**, 753 (1990).
 - [14] M. Creutz and A. Gocksch, Phys. Rev. Lett. **63**, 9 (1989).
 - [15] P. J. Channel and C. Scovel, Nonlinearity **3**, 231 (1990).
 - [16] H. Yoshida, Phys. Lett. **A150**, 262 (1990).
 - [17] B. Gladman, M. Duncan, and J. Candy, Celestial Mechanics and Dynamical Astronomy **52**, 221 (1991).
 - [18] M. Suzuki, Phys. Lett. **A146**, 319 (1990).
 - [19] M. Suzuki, J. Math. Phys. **32**, 400 (1991).
 - [20] B. Gladman, M. Duncan, and J. Candy, Celestial Mechanics and Dynamical Astronomy **52**, 221 (1991).
 - [21] J. C. Sexton and D. H. Weingarten, Nucl. Phys. **B380**, 665 (1992).
 - [22] H. Yoshida, Celestial Mechanics and Dynamical Astronomy **56**, 27 (1993).
 - [23] P. Hut, J. Makino, and S. McMillan, The Astrophysical Journal **443**, L93 (1995).
 - [24] A. D. Kennedy and P. Rossi, Nucl. Phys. **B327**, 782 (1989).
 - [25] S. Duane and J. B. Kogut, Phys. Rev. Lett. **55**, 2774 (1985).
 - [26] S. Duane and J. B. Kogut, Nucl. Phys. **B275**, 398 (1986).
 - [27] A. D. Kennedy and B. J. Pendleton, Nucl. Phys. **B607** [FS], 456 (2001), hep-lat/0008020.
 - [28] A. Horowitz, Nucl. Phys. **B280**[FS18], 510 (1987).
 - [29] A. M. Horowitz, Phys. Lett. **B268**, 247 (1991).
 - [30] J. Kuti, in *Computational Physics*, edited by R. D. Kenway and G. S. Pawley, Scottish Universities Summer School in Physics (Scottish Universities Summer School in Physics, 1987), pp. 311–378.
 - [31] M. Beccaria and G. Curci, Phys. Rev. **D49**, 2578 (1994), hep-lat/9307007.
 - [32] M. Beccaria, G. Curci, and L. Galli, Phys. Rev. **D49**, 2590 (1994), hep-lat/9307008.
 - [33] P. Langevin, Comptes Rendus **146**, 530 (1908).
 - [34] A. D. Fokker, Ann. Physik **43**, 810 (1914).
 - [35] M. Planck, Sitzber. Preuß. Akad. Wiss. p. 324 (1917).
 - [36] A. D. Kennedy and J. Kuti, Phys. Rev. Lett. **54**, 2473 (1985).
 - [37] G. Bhanot and A. D. Kennedy, Phys. Lett. **157B**, 70 (1985).
 - [38] L. Lin, K.-F. Liu, and J. H. Sloan, Phys. Rev. **D61**, 074505 (2000), hep-lat/9905033.
 - [39] T. D. Bakeyev and P. de Forcrand, Phys. Rev. **D63**, 054505 (2001), hep-lat/0008006.
 - [40] A. D. Kennedy and B. J. Pendleton, in *Lattice '90*, edited by U. M. Heller, A. D. Kennedy, and S. Sanielevici (1991), vol. B20 of *Nuclear Physics (Proceedings Supplements)*, pp. 118–121, talk presented at “Lattice '90,” Tallahassee.
 - [41] A. D. Kennedy and B. J. Pendleton, in *Lattice '99*, edited by M. Campostrini, S. Caracciolo, L. Cosmai, A. DiGiacomo, F. Rapuano, and P. Rossi (1999), vol. B83–84 of *Nuclear Physics (Proceedings Supplements)*, pp. 816–818, proceedings of the XVIIth International Symposium on Lattice Field Theory, Pisa, Italy, 29 June–3 July 1999, hep-lat/0001031.
 - [42] C. Reutenauer, *Free Lie Algebras*, vol. 7 of *London Mathematical Society Monographs, new series* (Oxford University Press, 1993), ISBN 0-19-853679-8.
 - [43] P. Hall, Proceedings of the London Mathematical Society **2**, 29 (1933).
 - [44] W. Magnus, Journal für die Reine und Angewandte Mathematik **177**, 105 (1937).
 - [45] M. Hall, Jr., Proceedings of the American Mathematical Society **1**, 575 (1950).
 - [46] E. Witt, Mathematische Annalen **177**, 152 (1937).

- [47] H. F. Baker, Proceedings of the London Mathematical Society **2**, 24 (1905).
- [48] J. E. Campbell, Proceedings of the London Mathematical Society **1**, 381 (1897).
- [49] J. E. Campbell, Proceedings of the London Mathematical Society **1**, 14 (1897).
- [50] F. Hausdorff, Leipziger Berichte **58**, 19 (1906).
- [51] J. Czyż, *Paradoxes of Measures and Dimensions Originating in Felix Hausdorff's Ideas* (World Scientific, Singapore, 1994), ISBN 9810201893.
- [52] V. S. Varadarajan, *Lie Groups, Lie Algebras, and their Representations*, Series in Modern Analysis (Prentice Hall, Englewood Cliffs, NJ, 1974), ISBN 0-13-535732-2.
- [53] E. Forest and R. D. Ruth, Physica **D43**, 105 (1990).